RNA-DEPENDENT DNA POLYMERASE FROM GEOBACILLUS STEAROTHERMOPHILUS

Field of the Invention

5   [001] The present invention relates to DNA and protein sequences encoding heat-stable polymerase enzymes, expression vector constructs for recombinant production of such enzymes, and methods of use for heat-stable polymerases. More specifically, the invention relates to a substantially pure thermostable RNA-directed DNA polymerase (*i.e.*, reverse

10   transcriptase) isolated from *Geobacillus stearothermophilus*. The invention also relates to the cloning and expression of the *G. stearothermophilus* RNA-directed DNA polymerase in *Escherichia coli*, to DNA molecules containing the cloned gene, and to hosts which express said genes.

15   ## Background of the Invention

        [002] Heat-stable enzymes are essential tools of molecular biology that have proven invaluable in DNA cloning, sequencing, and random mutagenesis. The most often used heat-stable polymerases are the DNA polymerases utilized in polymerase chain reactions (PCR) reactions. These

20   are often coupled with reverse transcriptases in RT-PCR, when an RNA molecule is used as a template to form a complementary DNA (cDNA) molecule, and that cDNA sequence is amplified by polymerase chain reaction (PCR). cDNA synthesis is most often done using reverse transcriptase enzymes of viral origin, with the reactions being performed at temperatures

25   below about 50°C, which is optimal for these enzymes. There are, however, distinct advantages to synthesizing cDNA at temperatures above 50°C, since higher temperatures melt the secondary structures that can form in the RNA template and block further processivity of the transcriptase enzyme. A

reverse transcriptase that remains stable and active at higher temperatures is especially useful for cDNA synthesis in combination reverse transcription/polymerase chain reaction (RT-PCR) reactions, as well as in other applications. Since higher temperatures can eliminate the secondary structures that may form in the RNA template, the length of the cDNA product can be extended. Higher temperatures also reduce the amount of non-specific annealing of PCR primers during cDNA synthesis, increasing specificity and amplification of cDNA. Higher temperatures can also melt the 3' end of a mismatched primer, inhibiting further synthesis from the primer and limiting incorporation of mismatched bases in the cDNA product.

[003] Retroviral reverse transcriptases generally have three potential enzymatic activities associated with them: an RNA-directed DNA polymerase, a DNA-directed DNA polymerase, and an RNAse H activity. Therefore, when retroviral RTs are used to copy RNA or DNA, an RNAse inhibitor must often be included in the reaction to minimize RNAse effects. Unfortunately, this can also inhibit the action of other enzymes. Furthermore, retroviral enzymes are typically most effective at temperatures at or below 50°C.

[004] To produce cDNA at higher temperatures, a DNA-dependent, DNA polymerase *Tth pol*, from the thermophilic bacterium *Thermus thermophilus*, has been used. Although it is not technically classified as a reverse transcriptase, it will reverse transcribe RNA at high temperatures in an RT-PCR reaction. Manganese chloride ($MnCl_2$) must be added to the reaction to boost efficiency, but this also reduces the fidelity of cDNA synthesis so that an added step is necessary to remove it prior to PCR amplification.

[005] An RNase H-deficient Avian Myeloblastosis Virus reverse transcriptase (AMV-RT) has been used in RT-PCR reactions, the RNase H-deficient enzyme being more thermostable than the native enzyme. This RT

2

does not degrade the RNA template, increasing the amount of full-length cDNA product that can be produced. cDNA synthesis with this enzyme is generally performed at 50°C, but larger amounts of the enzyme and substrate dNTPs are required for cDNA synthesis at higher temperatures, and synthesis from long RNA templates is often truncated, even at the higher temperature. AMV-RT also comprises two polypeptide chains (α and β), making it more difficult and expensive to produce as a recombinant product. When expressed in *E. coli*, for example, the end product is not appropriately modified to provide a fully active enzyme. An RT from Moloney Murine Leukemia Virus (MMLV) is also commercially available for use in reverse transcription reactions. Invitrogen's (Carlsbad, CA) SuperscriptII® RT is a point mutant of M-MLV-RT. According to the product literature, SuperscriptII® RT can be used at temperatures up to 50°C, and native M-MLV-RT can be used at temperatures up to 42°C.

[006] Although there are currently enzymes which can be used in RT-PCR and other similar types of reactions, there is still a significant need for an improved RT that remains active, accurate, and stable at high temperatures and can be used in a one-step reaction system for RT-PCR.

Summary of the Invention

[007] The present invention relates to a novel isolated polynucleotide sequence from the genome of *Bacillus stearothermophilus* (*Geobacillus stearothermophilus*) SEQ ID NO 1, and a novel amino acid sequence (SEQ ID NO 2) encoded by the polynucleotide sequence, the corresponding amino acid sequence comprising a heat-stable protein with reverse transcriptase activity. The invention provides a novel enzyme, *Tirt* (thermostable intron reverse transcriptase), which has reverse transcriptase activity and retains that activity at temperatures of up to about 75°C.

[008] The invention further relates to a method of using such a reverse transcriptase enzyme to facilitate DNA cloning, and at least one type of kit with prepared reagents to perform the necessary reactions for RT-PCR. The invention also provides a method of performing reverse transcription of an RNA sequence to form a cDNA sequence, and further provides a method for performing RT-PCR with the reverse transcription reaction being performed at temperatures elevated to about 75°C.

Brief Description of the Drawings

[009] FIG. 1 is a multiple amino acid sequence alignment comparison of the *Tirt* sequence with group II intron-encoded ORFs. The open reading frame (ORF) designated Bst 803-2065 encodes the *Tirt* protein and was cloned from the genome of *G. stearothermophilus* strain 10. The amino acid sequence of *Tirt* is compared with the amino acid sequence from three related ORFs encoded by group II introns from bacteria. These sequences include an ORF from *Bacillus halodurans* (ATCC accession no. NC002570.1), an RT-maturase protein from *Clostridium acetobutylicum* (ATCC accession no. NC003030), and a group II intron protein from *Pseudomonas alcaligenes* (ATCC accession no. U77945). Common to RTs, the *Tirt* sequence contains most of the highly conserved amino acids that fall into seven distinct domains (underlined sequences labeled I – VII). These seven conserved domains correspond to important structural regions proposed to be shared by all RTs. In addition, the Tirt sequence also contains most of the highly conserved amino acids contained in a region of the protein designated "X". This domain is associated with the maturase function of group II intron encoded proteins. The most conserved amino acids found in the "X" region of bacterial group II intron proteins (according to Zimmerly *et al. Nucleic Acids Res.* (2000) 29: 1238-1250) are shown above the alignment in italics.

[010] FIG. 2 is the polynucleotide sequence of the *tirt* ORF from *G. stearothermophilus.* The location of DNA sequences used to design the

4

primers Bst755, Bst1396, Bst2015, and Bst2198 are shown by the designated arrows. These primers were used to amplify the *tirt*-ORF for cloning.

[011] FIG. 3. is a photograph of a polyacrylamide protein gel (stained with Coomassie blue) illustrating over-expression of the Tirt fusion protein in *E. coli*. *E. coli* cells (BL21) containing the plasmid pTirt#16 were induced with IPTG to over-express the Tirt fusion protein via the T7 promoter system. Total protein extracts prepared from induced and non-induced cells were analyzed by polyacrylamide gel electrophoresis. A prominent protein band migrating at about 48 kD is apparent from induced cells (lane 4, indicated by arrow) but absent from non-induced cells (lane 3) and also absent from control cells containing just the plasmid vector alone (lane 1, uninduced and lane 2, induced cells). Most of the over-expressed protein band appears in the insoluble cell fraction (lane 8, soluble cell fraction versus lane 9, insoluble cell fraction). No 48 kD protein band appears in control cells containing just the expression plasmid (lane 6, soluble cell fraction and lane 7, insoluble cell fraction).

[012] FIG. 4. is a photograph of SDS-PAGE analysis of affinity column purification of the Tirt fusion protein. The soluble fraction from a cell extract expressing the Tirt fusion protein was loaded onto a nickel ion affinity column. A single-step elution of the polyhistidine tagged fusion protein yielded a partially purified fraction containing the 48 kD protein band (arrow). Column fractions 3-6 were analyzed by electrophoresis on a polyacrylamide protein gel stained with coomassie blue (lanes 6-9, respectively). The RT activity of each column fraction was stabilized by dialysis into buffer A (lanes 11-14). Lane 3 contains the insoluble cell fraction and lane 4 contains the soluble cell fraction that was loaded onto the column. Lane 5 contains the column flow-through. Lanes 1 and 2 contain a total cell extract from uninduced and induced cells respectively. Lane S contains a protein standard with the size (in kD) of each known protein indicated on the left.

[013] FIG. 5. is a schematic diagram showing the steps involved in the PERT assay used to detect RT activity from cell fractions. Column purified fractions of the Tirt protein serve as a source of RT activity and are added to a reaction mix containing BMV-RNA as a template. Using a specific primer (BMV-RT primer) the activity of the Tirt protein synthesizes a cDNA copy of the BMV-RNA template. A small region of the cDNA copy is then amplified (using specific primers) by the PCR to produce a final amplified product of 168 bp.

[014] FIG. 6. is a picture of a polyacrylamide gel illustrating the presence of RT activity in a purified preparation of the Tirt fusion protein. Purified preparations of the Tirt protein were added to the PERT assay as a source of RT. Various control reactions were also run with the PERT assay. The production of a 168 bp amplified DNA indicates RT activity. (A): Lane S, 100 bp molecular weight marker; lane 1, affinity purified column fraction #2; lane 2, column fraction #2 after dialysis in buffer A; lane 3, the same as lane 2 but supplemented with MMLV-RT; lane 4, MMLV-RT in RT buffer (serving as a positive control); lane 5, MMLV-RT but with one of the PCR primers absent from the reaction (a negative control). (B): Lane S, 100 bp molecular weight markers; lanes 1-3 contain column fractions # 2-4 respectively after dialysis in buffer A; lane 4, MMLV-RT in RT buffer (positive control); lane 5, MMLV-RT plus RNase A added to the reaction; lane 6, no exogenous (H_2O) source of RT added to the reaction; lane 7, column fraction #2 (dialyzed in buffer A) plus RNase A.

[015] FIG. 7. is a picture of a polyacrylamide gel illustrating the presence of heat stable RT activity associated with purified Tirt protein. The PERT assay was used to detect RT activity of the purified Tirt protein under various temperature conditions. (A) Column purified fractions of Tirt (dialyzed in buffer A) were pooled and incubated at the indicated temperatures for 15 minutes, then added to the PERT assay. Lane S, 100 bp molecular weight marker; lane 1, MMLV-RT with one PCR primer missing (negative control); lane 2, purified Tirt protein with no heat treatment; lane

6

3, purified Tirt protein heated to 65°C; lane 4, purified Tirt protein heated to 75°C; lane 5, commercial preparation of MMLV-RT with no heat treatment; lane 6, MMLV-RT heated to 65°C; lane 7, MMLV-RT heated to 75°C. (B) The PERT assay itself was run at three different temperatures with either the purified Tirt protein or MMLV-RT added as a source of RT. Lane S, 100 bp molecular weight marker; lane 1, a control reaction containing purified Tirt, but no RNA template was added (negative control); lanes 2-4, is the PERT assay containing the purified Tirt protein incubated at 37°C, 50°C and 68°C respectively; lanes 5-7, is the PERT assay containing MMLV-RT incubated at 37°C, 50°C, and 68°C respectively.

[016] Fig. 8. is a map of the pTirt#16 plasmid, for the over-expression and purification of the Tirt protein. (A) The restriction endonuclease map of plasmid pTirt#16 indicates the location of the genomic DNA (hatched rectangle) cloned from *G. stearothermophilus* that contains the Tirt ORF. The Tirt ORF was cloned adjacent to a T7 promoter and in frame with a poly-histidine ([His]$_6$) tag to allow over-expression of the Tirt protein and simplify purification. (B) A partial DNA sequence of the pTirt#16 plasmid shows the junction region where the Tirt ORF is fused in frame with the poly-histidine tag element (underlined sequence) down stream of the T7 promoter, creating a 35 amino acid extension at the N-terminus of the Tirt ORF that includes six consecutive histidine amino acids.

Detailed Description

[017] The present invention seeks to overcome the shortcomings of the prior art by providing a thermostable RNA-directed DNA polymerase as found in the thermophilic bacterium *Geobacillus stearothermophilus* (also known as *Bacillus stearothermophilus*). *Geobacillus stearothermophilus* is an

7

organism isolated from the formation waters of Russian oilfields. A DNA-dependent DNA polymerase from this organism has previously been described. The inventors describe here the discovery of a functional RNA-dependent DNA polymerase enzyme encoded within the genome of *G.*

5   *stearothermophilus.* While investigating putative coding regions within the sequence data, the inventors discovered a sequence having characteristics similar to those of type II introns, which have reverse transcriptase activity. In methods to be described herein, the inventors isolated DNA encoding a protein they have designated as Tirt (for "thermostable intron reverse

10   transcriptase"), constructed an expression vector for its overexpression, and demonstrated that it has reverse transcriptase activity even after being heated to about 75°C.

[018] The DNA sequence shown in Fig. 2 is from the incomplete genomic DNA sequence of *Bacillus (Geobacillus) stearothermophilus* strain 10

15   obtained from the Genome Sequencing Project, Advanced Center for Genome Technology, University of Oklahoma (Experimental Program to Stimulate Competitive Research Grant #EPS-9550478). The inventors have discovered that the sequence encodes a 420-amino acid ORF containing the *Tirt* protein.

[019] Retroelements are genetic elements that code for a reverse

20   transcriptase and employ the process of reverse transcription in some stage of their replication or mobility. A large variety of these retroelements are found in nearly every type of eukaryotic organism. They include some RNA viruses, DNA viruses, transposons, introns, and even mitochondrial plasmids. Bacteria also contain RT-encoding genetic elements (retro-

25   elements) that fall into two basic types. The group II introns, found in a variety of bacteria, contain a reverse transcriptase region as part of the intron-encoded ORF. A retron, on the other hand, produces an unusual satellite DNA called msDNA. The inventors compared the amino acid sequence from ten different retron RTs with the amino acid sequence from

30   five different bacterial group II intron ORFs by multiple sequence alignment

8

(Clustal W alignment). They then used the multiple sequence alignment to generate a consensus amino acid sequence for RTs found in bacteria. The bacterial consensus sequence was used as the query sequence in a BLAST search of both the GenBank database as well as the bacterial genomes database that contains both completed, as well as, unfinished bacterial species. These searches revealed several ORFs with similarity to the consensus sequence that had not been previously described. This included a 420-amino acid ORF from the unfinished genome sequence of *G. stearothermophilus*.

[020] The amino acid sequence of this *G. stearothermophilus* ORF was further analyzed by comparison with other RTs and found to be strongly similar to group II intron ORFs from both bacteria and mitochondria (Fig. 1). The amino acid sequences of RTs are generally highly variable. However, multiple amino acid sequence alignments demonstrate the presence of a few highly conserved amino acids shared among all RTs. These conserved amino acids fall into seven domains (designated I – VII, Fig. 1) that correspond to conserved secondary structures within the folded RT protein. A short distance beyond domain VII is an additional conserved region designated domain "X", that is found only in group II intron encoded proteins. Domain X is associated with the maturase function found in group II intron ORFs. Although domain X is not well conserved among bacterial group II intron proteins, the *G. stearothermophilus* ORF appears to contain most of the conserved amino acids of domain X shared among bacterial group II introns (Fig. 1). Thus, the *G. stearothermophilus* ORF appears to have most of the highly conserved amino acids present in both the RT region and the maturase region (or domain X) of group II intron proteins (Fig. 1). For this reason the *G. stearothermophilus* ORF is clearly not a retron type RT and was thus designated *tirt* for thermostable intron reverse transcriptase. Some proteins encoded by group II introns also contain a third, zinc finger domain that imparts an endonuclease activity on this multifunctional protein.

However, based on sequence comparisons, this endonuclease domain appears to be absent from the *tirt* ORF.

[021] As used herein, the term "protein" is intended to include mimetics and the term "amino acid" is intended to include L-form, D-form, and modified amino acids. These substitutions may be made by someone of skill in the art, using the known structural similarities between the molecules. The amino acid sequence is also intended to include any peptide or protein sequence that may include additional amino acids either N-terminal or C-terminal to the listed sequence, or both. The term "Tirt protein" is intended to include variants or biologically active fragments of the polypeptide.

[022] It is well known in the art that a single amino acid may be encoded by more than one nucleotide codon, and that the nucleotide sequence may be modified to produce an alternate nucleotide sequence that encodes the same peptide. Therefore, alternate embodiments of the present invention include alternate DNA sequences encoding peptides containing the amino acid sequences as previously described. DNA sequences encoding peptides containing the claimed amino acid sequence include DNA sequences which encode any combination of the claimed sequence and any other amino acids located N-terminal or C-terminal to the claimed amino acid sequence.

[023] It is to be understood that amino acid and nucleic acid sequences may include additional residues, particularly N- or C-terminal amino acids or 5' or 3' nucleotide sequences, and still be essentially as set forth in the sequences disclosed herein, as long as the sequence confers heat-stable RNA-dependent DNA polymerase activity upon the polypeptide or protein moiety of the expressed protein.

[024] Additional nucleic acid bases may be added either 5' or 3' to the Tirt ORF, and may be combined with other DNA sequences, such as

promoters, polyadenylation signals, additional restriction enzyme sites, multiple cloning sites, other coding segments, and the like. Therefore, overall length of such a polynucleotide may vary considerably. In a method described by the present invention, a nucleotide sequence as shown in Fig. 2

5   is inserted into a protein expression vector to produce a protein which can be used to synthesize a DNA copy of an RNA molecule. The DNA can then be amplified to form multiple copies, at temperatures elevated to about 68 to about 75 degrees Celsius.

[025] "Control sequences" are those DNA sequences that are necessary
10   for the expression of a protein from a polynucleotide sequence containing such a sequence, operably linked to the polynucleotide sequence encoding the protein. These sequences include prokaryotic sequences such as, for example, promoters, operators, and ribosome binding sites, and eukaryotic sequences such as, for example, promoters, enhancers, and polyadenylation
15   signals. "Expression systems" are DNA sequences (such as, for example, plasmids) appropriate for expression of a target protein in a particular host cell, these sequences comprising appropriate control sequences for protein expression in the host cell operably linked to the polynucleotide sequence encoding the target protein.

20   [026] It is to be understood that a "variant" of a polypeptide is not completely identical to the native protein. A variant Tirt protein, for example, can be obtained by altering the amino acid sequence by insertion, deletion or substitution of one or more amino acids. The amino acid sequence of the protein can be modified, for example, by substitution to
25   create a polypeptide having substantially the same or improved qualities as compared to the native polypeptide. The substitution may be a conserved substitution. A "conserved substitution" is a substitution of an amino acid with another amino acid having a side chain that is similar in polar/nonpolar nature, charge, or size. The 20 essential amino acids can be

11

grouped as those having nonpolar side chains (alanine, valine, leucine, isoleucine, proline, phenylalanine, and tryptophan), uncharged polar side chains (methionine, glycine, serine, threonine, cystine, tyrosine, asparagine and glutamine), acidic side chains (aspartate and glutamate), and basic side

5   chains (lysine, arginine, and histidine). Conserved substitutions might include, for example, Asp to Glu, Asn, or Gln; His to Lys, Arg or Phe; Asn to Gln, Asp or Glu; and Ser to Cys, Thr or Gly. Alanine, for example, is often used to make conserved substitutions.

[027] To those of skill in the art, variant polypeptides can be obtained

10   by substituting a first amino acid for a second amino acid at one or more positions in the polypeptide structure in order to affect biological activity. Amino acid substitutions may, for example, induce conformational changes in a polypeptide that result in increased biological activity.

[028] Those of skill in the art may also make substitutions in the

15   amino acid sequence based on the hydrophilicity index or hydropathic index of the amino acids. A variant amino acid molecule of the present invention, therefore, has less than one hundred percent, but at least about fifty percent, and preferably at least about eighty to about ninety percent amino acid sequence homology or identity to the amino acid sequence of a

20   polypeptide comprising SEQ ID NO 2, or a polypeptide encoded by SEQ ID NO 1. Therefore, the amino acid sequence of the variant Tirt protein corresponds essentially to the native Tirt protein amino acid sequence. As used herein, "corresponds essentially to" refers to a polypeptide sequence that will elicit a similar biological and enzymatic activity to that generated by

25   a Tirt protein comprising SEQ ID NO 2, such activity being at least about 70 percent that of the native Tirt protein, and more preferably greater than 100 percent of the activity of the native Tirt protein.

[029] A variant of the Tirt protein may include amino acid residues not present in a corresponding Tirt protein comprising SEQ ID NO 2, or may

include deletions relative to the Tirt protein comprising SEQ ID NO 2. A variant may also be a truncated "fragment," as compared to the corresponding protein comprising SEQ ID NO 2, the fragment being only a portion of the full-length protein.

[030] In a preferred embodiment of the present invention, a protein expression vector is genetically engineered to incorporate a DNA sequence encoding the Tirt ORF and appropriate control sequences comprising, for example, transcriptional and translational sequences such as promoter and polyadenylation sequences, to produce a functional Tirt protein.

[031] The isolated polynucleotide and protein of the present invention can be used in a variety of applications, including, but not limited to assays to confirm the presence of viral infection in tissue samples, preparation of cDNA copies of isolated or cellular RNAs, and real-time or standard RT-PCR for genetic analysis. These and other uses known to, or developed by, those of skill in the art are made possible by the discovery of the thermostable intron reverse transcriptase of the present invention.

[032] Expression vectors may be chosen from among those readily available for prokaryotic or eukaryotic expression systems. Expression system vectors, which incorporate the necessary regulatory elements for protein expression, as well as restriction endonuclease sites that facilitate cloning of the desired sequences into the vector, are known to those of skill in the art. A number of these expression vectors are commercially available. In one preferred embodiment of the present invention, the expression vector is pET28 (Novagen, Madison, WI).

[033] An expression vector host cell system can be chosen from among a number of such systems that are known to those of skill in the art. In one embodiment of the invention, the protein can be expressed in *E. coli.* In alternate embodiments of the present invention, the enzyme may be

13

expressed and purified using other bacterial expression systems, viral expression systems, eukaryotic expression systems, or cell-free expression systems. Cellular hosts used by those of skill in the art for expression of various proteins include, but are not limited to, *Bacillus subtilis*, yeast such

5 as *Saccharomyces cerevisiae*, *Saccharomyces carlsbergenesis*, *Saccharomyces pombe*, and *Pichia pastoris*, as well as mammalian cells such as 3T3, HeLa, and Vero. The expression vector chosen by one of skill in the art will include promoter elements and other regulatory elements appropriate for the host cell or cell-free system in which the recombinant DNA sequence

10 encoding the enzyme will be expressed. In mammalian expression systems, for example, suitable expression vectors can include DNA plasmids, DNA viruses, and RNA viruses. In bacterial expression systems, suitable vectors can include plasmid DNA and bacteriophage vectors.

[034] One group of vectors that can be used to express and facilitate

15 purification of the protein include those vectors that encode the polyhistidine (6xHis) sequence and an epitope tag to allow rapid purification of the fusion protein with a nickel-chelating resin, along with protein detection with specific antibodies to detect the presence of the secreted protein. An example of such a vector for expression in mammalian cells is the

20 pcDNA3.1/V5-His-TOPO eukaryotic expression vector (Invitrogen). In this vector, the fusion protein can be expressed at high levels under the control of a strong cytomegalovirus (CMV) promoter. A C-terminal polyhistidine (6xHis) tag enables fusion protein purification using nickel-chelating resin. Secreted protein produced by this vector can be detected using an anti-His

25 (C-term) antibody.

[035] Since Tirt is a bacterial protein, bacterial expression systems are particularly suited for expression of the Tirt protein as described by the present invention. Such systems include, for example, the pMAL system (New England Biolabs, Beverly, MA) which utilizes a maltose binding protein

14

fusion to facilitate purification, and the Impact-CN Protein Fusion and Purification System (New England Biolabs).

[036] A baculovirus expression system can be used for production of a target protein such as the enzyme of the present invention. A commonly used baculovirus is AcMNPV. Cloning of the target protein DNA can be accomplished by using homologous recombination. The target protein DNA sequence is cloned into a transfer vector containing a baculovirus promoter flanked by baculovirus DNA, particularly DNA from the polyhedrin gene. This DNA is transfected into insect cells, where homologous recombination occurs to insert the target protein into the genome of the parent virus. Recombinants are identified by altered plaque morphology.

[037] Proteins as described above can also be produced in the method of the present invention by mammalian viral expression systems. The Sindbis viral expression system, for example, can be used to express proteins at high levels. Sindbis vectors have been described, for example, in U.S. Pat. No. 5,091,309 (Schlesinger *et al.*), incorporated herein by reference. Sindbis expression vectors, such as pSinHis (Invitrogen, Carlsbad, Calif.) can be used to express the Tirt protein under the direction of the subgenomic promoter PSG. In vitro transcribed RNA molecules encoding the fusion protein and the Sindbis proteins required for in vivo RNA amplification can be electroporated into baby hamster kidney (BHK) cells using methods known to those of skill in the art. Alternatively, the RNA encoding the Tirt protein and Sindbis proteins required for in vivo RNA amplification can be cotransfected with helper RNA that permits the production of recombinant viral particles. Viral particles containing genetic material encoding the fusion protein can then be used to infect cells of a wide variety of cell types, including mammalian, avian, reptilian, and Drosophila. Fusion protein expressed from the pSinHis (Invitrogen) vector can be detected with antibody to an Anti-Xpress™ epitope encoded by the vector sequence. The pSinHis

vector also includes a polyhistidine tag which provides a binding site for metal-chelating resins to facilitate purification of the expressed fusion protein. Furthermore, an enterokinase cleavage site located between the histidine tag and the fusion protein allows the histidine tag to be enzymatically removed following purification.

[038] An ecdysone-inducible mammalian expression system (Invitrogen, Carlsbad, Calif.) can also be used to express a target protein. Vectors used in the ecdysone-inducible mammalian expression system can be organized to produce the target protein by expressing the target protein from the expression cassette. With the ecdysone-inducible system, higher levels of protein production can be achieved by use of the insect hormone 20-OH ecdysone to activate gene expression via the ecdysone receptor. An inducible expression plasmid provides a multiple cloning site, into which the nucleotide sequence of the Tirt protein can be ligated. The expression vector contains ecdysone response elements upstream of the promoter (a minimal heat shock promoter) and the multiple cloning site. Cotransfection of a second plasmid, pVgRXR (Invitrogen), provides the receptor subunits to make the cell responsive to the steroid hormone ecdysone analog, ponasterone A. A control expression plasmid containing the lacZ gene can be cotransfected with pVgRXR to provide a marker for transfected cells. Upon induction with ponasterone A, the control plasmid expresses ß-galactosidase. Cotransfection of the inducible expression construct and pVgRXR into the mammalian cell of choice can be accomplished by any of the standard means known to those of skill in the art. These include, for example, calcium phosphate transfection, lipid-mediated transfection, and electroporation. Levels of expression of the fusion protein in this system can be varied according to the concentration and length of exposure to ponasterone. Stable cell lines that constitutively express the Tirt protein can be established using Zeocin.TM. (Invitrogen), a bleomycin/phleomycin-type antibiotic isolated from *Streptomyces*, and neomycin or hygromycin.

[039] Yeast host cells, such as *Pichia pastoris*, can also be used for the production of the Tirt protein. Expression of heterologous proteins from plasmids transformed into *Pichia* has previously been described by Sreekrishna, *et al.* (U.S. Pat. No. 5,002,876, incorporated herein by reference). Vectors for expression in *Pichia* of a Tirt protein are commercially available as part of a Pichia Expression Kit (Invitrogen, Carlsbad, Calif.). *Pichia pastoris* is a methylotrophic yeast, which produces large amounts of alcohol oxidase to avoid the toxicity of hydrogen peroxide produced as a result of methanol metabolism. Alcohol oxidase gene expression is tightly regulated by the AOX1 and AOX2 promoters. In *Pichia* expression vectors, high levels of expression are produced under the control of these promoters. Ohi, *et al.* (U.S. Pat. No. 5,683,893, incorporated herein by reference) have previously described a mutant AOX2 promoter capable of producing enhanced expression levels.

[040] PCR primers were designed based on the DNA sequence recovered from a BLAST search of the unfinished genome sequence of *G. stearothermophilus* from the "BLAST with bacterial genomes" web page at the National Center for Biotech Information (www.ncbi.nlm.nih-gov/blast) (see Fig. 1). The program Primer3 (available at the web site: www-genome.wi. mit.edu/cgi-bin/primer/primer3) was used to aid in design of primers. The DNA sequence of plasmid pTirt#16 and other constructs was determined by BigDye terminator cycle sequencing using an ABI 327 automated DNA sequencer at the sequencing service (Molecular Biology Core facility) provided by the Quillen College of Medicine at East Tennessee State University. Multiple amino acid sequence alignments of the tirt ORF with other known intron ORFs was done using Clustal W available from the Baylor College of Medicine at http://searchlauncher.bc.tmc.edu/).

[041] Protocols for performing reverse transcription, RT-PCR, real-time RT-PCR, real-time relative RT-PCR, construction of cDNA libraries, and competitive reverse transcriptase PCR analysis of cellular genes are known

to those of skill in the art. One such protocol for competitive reverse transcriptase PCR analysis is described by Waha, *et al.* in <u>Brain Pathology</u>, Vol. 8 (1998), pages 13 to 18. Protocols for detecting the presence of viral infection include, for example, those described by Henrickson, *et al.*, in

5   United States Patents No. 5,744,299 (parainfluenza virus type 1) and 6,014,664 (parainfluenza virus, respiratory syncytial virus, and influenza virus). More recently, detection of the Severe Acute Respiratory Syndrome (SARS) Coronavirus by real-time nested PCR following RT-PCR. (Jiang, *et al*, <u>Clin. Infectious Disease</u> 2004: 38 (15 Jan.) p. 293-296.) The enzyme of the

10   present invention provides a tool for performing these protocols at higher temperatures. It is known, for example, that in the construction of cDNA libraries or in cDNA labeling, that oligo(dT) primers can be used to insure that poly(A) mRNAs are reverse transcribed, and that short random oligonucleotide primers may also be used for reverse transcription. Those of

15   skill in the art, however, also are aware that RT reactions are often performed at approximately 42°C in order to avoid inactivating the reverse transcriptase needed for the reaction to be catalyzed. The present invention provides an alternate enzyme for reverse transcription reactions, such as those listed above, that can be used at temperatures significantly higher

20   than those used currently in conventional reaction systems.

[042] The invention also provides at least one kit for performing, for example, reverse transcriptase, RT-PCR, real-time RT-PCR, competitive RT-PCR, or other protocols which rely on the presence of a reverse transcriptase, these protocols being performed at temperatures of up to about 68 to 75

25   degrees Celsius. Such a kit may comprise, for example, at least one reaction buffer (e.g., 50 mM Tris-pH 8.3, 100 mM KCl, and 10 mM dithiothreitol), an RNase inhibitor (to a final concentration of 10 units), NP-40 (to a final concentration of 0.17%), dNTP mix (dGTP, dCTP, dUTP, and dATP to a final concentration of 0.8 mM each), Tirt protein, Taq polymerase, BMV-RNA (as a

30   positive control to a final concentration of 50 ng), RT primer (as a positive control primer for RT, to a final concentration of 0.02 μM), and BMV-PCR1

primer with BMV-PCR2 primer (as positive control primers for PCR, to a final concentration of 1 µM).

[043] The present invention provides the isolated polynucleotide encoding the amino acid sequence of SEQ ID NO: 2, which can be used in production of the reverse transcriptase of the invention and the variant thereof. Furthermore, nucleic acids which hybridize with a nucleic acid encoding the amino acid sequence of SEQ ID NO: 2 under stringent conditions and encode a polypeptide having a similar reverse transcriptase activity to that of a polypeptide comprising SEQ ID NO: 2 are also included as embodiments of the present invention.

[044] The term "stringent conditions", as used herein, means conditions in which non-specific hybridization will not generally occur. Hybridization under such conditions can be performed based on the description provided in Molecular Cloning : A Laboratory Manual 2nd ed., published by cold Spring Harbor Laboratory in 1989, edited by T. Maniatis et al. For example, stringent conditions include incubation with a probe in 6×SSC containing 0.5% SDS, 5×Denhardt's solution and 100 micrograms /ml salmon sperm DNA at 60°C.

[045] The invention will be further described by means of the following non-limiting examples.

Example 1: Cloning and Expression of the *tirt* Gene

[046] *Geobacillus stearothermophilus* strain 10 was a kind gift from Dr. Bruce Roe, University of Oklahoma, and was used for cloning the *tirt* gene. Cultures were grown in trypticase soy agar plates at 55°C. The plasmid pUC18 was used for routine subcloning of DNA fragments. The plasmid pET28a and the *E. coli* strain BL21(DE3) were from Novagen (Madison, WI) and were used for heterologous expression of the Tirt protein in *E. coli.*

[047] While multiple sets of oligonucleotide primers were synthesized to amplify the *tirt* ORF, the inventors found that two pairs of primers could be used to successfully amplifying the *tirt* gene from the chromosome of *G.*

19

stearothermophilus. The first primer pair, designated Bst755 and Bst2015 was used to specifically amplify most, but not all, of the *tirt* ORF via a simple PCR protocol (Fig. 2). A second primer pair designated Bst1396 and Bst2198 was used to amplify a region that includes the last 16 amino acids

5   at the C-terminus of the *tirt* ORF (Fig. 2).

[048] The primers used to amplify the *tirt* gene from the genome of *G. stearothermophilus* strain 10 comprise SEQ ID NO 4: AGACAA<u>CATATG</u>CGGCAAGACCTGAATCTCAT-3' (with the underlined sequence indicating an *Nde*I restriction site for cloning into the pET28a

10   expression vector); SEQ ID NO 5: 5'-AAT<u>GGATCC</u>GCTGGCGAACATCCTTCTC-3' (with the underlined sequence indicating a *Bam*HI restriction site): SEQ ID NO 6: 5'-ATTA<u>CTGCAG</u>AGCGGTCCAGTAGGTTTTG-3' (with the underlined sequence indicating a *Pst*I restriction site); and SEQ ID NO 7: 5'-

15   ACTC<u>AAGCTT</u>GAGAAGGGCTTGACGTTCATG-3' (with the underlined sequence indicating a *Hind*III restriction site for cloning into the pET28a expression vector.).

[049] Amplification of the *tirt* gene was done in two stages (Fig. 2) using a single colony of *G. stearothermophilus* as a source of template. A

20   single colony from an overnight plate culture was suspended in 10 µl of water. One µl of this cell suspension was added to a 50 µl PCR reaction mix containing 1.5 mM MgCl2, Taq polymerase buffer (Promega, Madison, WI), 0.2 µM each dNTP, 0.5 µM each primer, and 2 units of Taq polymerase. The reaction was amplified using the following conditions: one cycle at 95°C for

25   2 minutes, 30 cycles at 95°C for one minute, 50°C for 2 minutes, and 72°C for 2 minutes. Amplified DNAs were purified by gel electrophoresis, digested with the appropriate restriction endonuclease and ligated into either pUC18 or directly into the pET28a expression vector. First, a 1.26 kilobase pair (Kb) amplified DNA product produced by the first primer pair (SEQ ID NO 4 and

30   SEQ ID NO 6) was ligated into the expression plasmid pET28a. This

produced an in-frame fusion between the polyhistidine tag element found in the expression plasmid and the *tirt* ORF. To capture the remaining 16 amino acids at the C-terminus of Tirt, a naturally occurring *Eco*RI site (within the *tirt* ORF, Fig. 2) was used to splice the 3 prime end of the PCR product

5   produced by the second primer pair (SEQ ID NO 5 and SEQ ID NO 7) to the first amplified DNA to yield an expression plasmid, pTirt#16, containing the entire predicted ORF of the Tirt protein.

Example 2: Expression of Tirt protein

10   [050] The plasmid pTirt#16 was used to express the Tirt protein in *E. coli* by induction of the T-7 promoter system with IPTG. Briefly, protein expression in *E. coli* was achieved using the T7 RNA polymerase system and the pET28a expression vector (Novagen, Madison, WI). Since the upstream primer (SEQ ID NO 4) used to amplify the *tirt* gene contained an *Nde*I

15   restriction site, the amplified DNA containing the *tirt* ORF could be ligated into the *Nde*I restriction site of the expression vector. This produced an in-frame fusion between the polyhistidine tag element in the pET28a vector and the tirt ORF.

[051] The Tirt fusion protein was expressed in *E. coli* strain BL21

20   lysogenic for μDE3. Briefly, cells of *E. coli* strain BL21(DE3) transformed with the plasmid pTirt#16 (containing the *tirt* fusion construct) were induced by addition of IPTG (1mM). After 3 hours of induction, cells from a 100 ml culture were harvested and resuspended in binding buffer (1X) for nickel ion column purification (Novagen). A cell extract containing the Tirt fusion

25   protein was prepared by incubating the cell suspension in fresh lysozyme (1 mg/ml), followed by three cycles of quick freeze-thaw (10 minutes at –80°C, followed by 10 minutes at 37°C), followed by sonication. Centrifugation (15,000 X g) and filtration (0.45μ filter) produced a cleared, crude protein preparation. The cleared extract was then loaded onto a prepared Nickel ion

column (His-bind column, Novagen) and purified fractions collected according to manufacturer's instructions (Novagen).

[052] Soluble and insoluble protein fractions were compared by resuspending cells (from a 50 ml induced culture) in one tenth volume of
5   buffer (50 mM Tris-HCl pH 8.0, 2 mM EDTA) containing lysozyme (100 μg/ml) plus 1% triton X-100. After incubation at 30°C for 15 minutes and sonication, the cell extract was centrifuged at 12,000 X g for 15 minutes. The supernatant was mixed with an equal volume of SDS sample buffer and this served as the soluble protein fraction for protein gels. The pellet of cell
10 debris was mixed with SDS sample buffer and this served as the insoluble protein fraction for protein gels.

[053] Only cell extracts from IPTG-induced cultures containing pTirt#16 showed a prominent protein band when analyzed by polyacrylamide gel electrophoresis (SDS-PAGE). The expressed protein appeared to be about
15 48 kilodaltons (kD) in size, according to its migration during electrophoresis in a polyacrylamide protein gel (Fig. 3, lane 4). This was approximately the size expected (52 kD) for the predicted fusion construction of Tirt in plasmid pTirt#16. In addition, Western blot analysis with a specific antibody probe confirmed the presence of a polyhistidine tag in the 48 kD protein band.
20 Most of the expressed fusion protein fractionated into the insoluble cell debris after high speed centrifugation of the cell extract (Fig. 3, lane 9). However, some of the 48 kD fusion protein also appeared in the soluble cell fraction (Fig. 3, lane 8) and the inventors suspected that his fraction might have detectable RT activity. The 48 kD fusion protein was purified from the
25 soluble cell fraction by nickel ion affinity chromatography. A single-step elution of the polyhistidine-tagged fusion protein yielded a partially purified fraction containing predominately a 48 kD protein band (Fig. 4, lanes 6-9, indicated by arrow). Each eluted column fraction was dialyzed and concentrated into a new buffer (buffer A) to stabilize the purified protein (Fig.
30 4, lanes 11-14). The polyhistidine tag at the N-terminus of the purified Tirt protein was not removed because this short extension of the protein was not

expected to affect the RT activity of the fusion protein, since it does not affect the expression of mammalian viral RTs and other recombinant eukaryotic RTs in *E. coli* when similar technology is used for their expression.

Example 3: Demonstration of Tirt's Reverse Transcriptase Activity

[054] Column fractions containing the purified Tirt fusion protein were pooled, dialyzed into buffer A (50mM Tris-pH 7.5, 1mM EDTA, 1mM DTT, and 10% glycerol) using a microcon 30 membrane concentrator (Amicon, Beverly, MA), and then used to assay for RT activity. A highly sensitive product enhanced reverse transcriptase (PERT) assay was used to detect RT activity. The assay required the reverse transcription of a Brome Mosaic virus (BMV) RNA template to produce a small cDNA that was then further amplified by PCR (Fig. 5). Briefly, the assay was performed by first assembling the PCR amplification reaction mix in the bottom of a 0.2 ml tube containing: $MgCl_2$-free PCR buffer, 1X (Promega, Madison, WI), 1µM each BMV-PCR1 primer (5'-CGTGGTTGACACGCAGACCTCTTAC-3') and BMV-PCR2 primer (5'-TCAACACTGTACGGCACCCGCATTC-3'), 0.8 mM each dNTP, and Taq polymerase (Promega). The RT reaction mix was then assembled on top after sealing the lower PCR reaction mix with a layer of wax using an Ampliwax pellet (PCR-Gem 50, Applied Biosystems, Roche). The RT reaction mix contained: RT buffer (50mM Tris-pH 8.3, 75mM KCl, and 10mM DTT), 2.5 mM $MgCl_2$, 0.17% NP-40, 10 units of RNasin (Promega, Madison, WI), 0.8mM each dNTP, 0.02µM RT primer (5'-GGTCTCTTTTAGAGATTTACAGTG-3'), 100ng of Brome Mosaic Virus (BMV) RNA (Promega), and a source of RT. The source of RT added to the reaction was either the purified Tirt fusion protein described above or commercially available Moloney Murine Leukemia Virus RT (MMLV-RT) (2 units). The reaction tube was then placed in a thermocycler under the following conditions: 1 cycle at 37°C for 1 hour (reverse transcription); 1 cycle at 94°C for 1 minute; 30 cycles at 94°C for 15 seconds, 56°C for 15 seconds, and 72°C for 15 seconds (amplification); and finally 72°C for 5 minutes.

Amplified DNA was detected by electrophoresis of the reaction mix on a 5% polyacrylamide gel followed by staining with ethidium bromide.

[055] The presence of a 168 base pair (bp) PCR product in a DNA gel, following electrophoresis, indicated the presence of RT activity. The assay was run on both crude cell extracts, as well as purified column fractions. No RT activity was detected in the crude cell extracts tested. Furthermore, no RT activity was detected in the purified fractions eluted from the nickel ion affinity column used to purify the fusion protein (Fig. 6A, lane 1). However, when the eluted column fractions were dialyzed and concentrated into buffer A, RT activity was detected in some of the column fractions (Fig. 6B, lanes 1-3), indicating the presence of inhibitors of RT activity in the fractions prior to dialysis. As an additional indication that inhibitors of RT activity are present in the column fractions prior to dialysis, when commercially prepared MMLV-RT was added to the PERT assay, as expected, a 168 bp DNA was produced indicating RT activity (Fig. 6A, lane 4). However, when MMLV-RT was mixed with column fraction #7, no RT activity was detected.

[056] When one of the PCR primers was omitted from the PERT assay, no DNA product was produced (Fig. 6A: lane 4, two primers; lane 5, one primer). This indicated that the 168 bp DNA product was the result of specific amplification of the reverse transcribed cDNA and not some other process. In another control reaction, RNase was added to the fraction to be tested for RT activity. When RNase was mixed with dialyzed column fraction number 2, no DNA product was observed (Fig. 6B lane 1, without RNase; lane 7, RNase added), indicating that the 168 bp DNA was amplified from a cDNA that was reversed transcribed from the BMV RNA template present in the assay and not from contaminating DNA carried over from previous assay reactions. A third control reaction contained only water as the sample extract to be tested for RT activity (Fig. 6B, lane 6). As expected, no DNA product was produced. This confirmed that an exogenous source of RT added to the assay reaction, and not the Taq polymerase present in the

assay, was responsible for cDNA formation (and thus the amplification of the 168 bp DNA).

Example 4: Demonstration of Heat Stability of Tirt

[057] The column purified Tirt fusion protein (as described above) was added (15 µl) to a microfuge tube and heated in a water bath at either 65°C or 75°C for 15 minutes. After heat treatment, the tube was centrifuged and 7.5 µl of the supernatant was added directly to the PERT assay reaction. MMLV-RT diluted in 1X RT buffer (5 units) was treated in a similar fashion.

[058] To run the PERT assay at different temperatures, the reverse transcription reaction was incubated separately at either 37°C, 50°C or 68°C. After 1 hour of incubation the reaction mix was immediately added to the top of the PCR reaction mix, sealed with a layer of wax, and processed as described in Example 3.

[059] The column-purified Tirt fusion protein was not heated (Fig. 7A, lane 2), heated at 65°C for 15 minutes (Fig. 7A, lane 3), or heated at 75°C for 15 minutes (Fig. 7A, lane 4). After heat treatment the purified fraction was tested for RT activity using the PERT assay.   As a control, the same procedure was followed using commercially prepared Moloney Murine Leukemia Virus reverse transcriptase (M-MLV-RT). The purified Tirt protein was not reduced in RT activity even after exposure to temperatures of 75°C (Fig. 7A, lane 4).   In contrast, no RT activity was detected after heat treatment (both at 65°C and 75°C) of the mesophilic MMLV-RT(Fig. 7A, lanes 6 and 7).

[058] The purified Tirt protein was also tested for RT activity by running the PERT assay at three different temperatures; 37°C, 50°C, and 68°C (Fig. 7B). Again, the purified Tirt protein was not affected in its ability to synthesize cDNA by reverse transcription even at the highest temperature tested, 68°C (Fig. 7B, lanes 2 and 4).   By contrast, cDNA production appeared to be greatly reduced, at least at the highest temperature of 68°C, for the mesophilic mammalian RT (compare Fig. 7B, lane 5 versus lane 7).

25

Plasmid pTirt#16 (comprising vector pET28a from Escherichia coli with Tirt coding sequence from Geobacillus stearothermophilus inserted) was deposited, prior to the date of filing of this application, with the American Type Culture Collection, Manassas, Virginia, USA, under the terms

5   of the Budapest Treaty. A computer readable sequence listing accompanies this application, and the sequence listing information recorded in computer readable form is identical to the written sequence listing included in the application papers.